

National Cancer Institute

External Data Standards Review

Exhibit 1. Standard Type Table

Standard Type	Standard Name	Standard Content	Recommendation
Common Demographic/Information Processing and Code Sets			
Address	Federal Geographic Data Committee (FGDC) Address Data Content Standard (Draft)	Address information for a physical location.	C
	FIPS 5-2, Codes for Identification of the States, District of Columbia, and Outlying Areas of the United States, and Associated Areas	State names	A
	FIPS 10-4, Countries, Dependencies, Areas of Special Sovereignty, and their Principal Administrative Divisions	Country names	A
	ISO 3166-1, Country Codes	Country codes	A
	ISO 11180:1993, Postal Addressing	Address format	A
	Universal Postal Union	Address format, state codes, country codes	A
	U.S. Postal Service Postal Addressing Standards	Address format, state codes, street suffixes, secondary unit designators	B
Language	ISO 639, Codes for representation of language	Language codes	A
Race and Ethnicity	Office of Management and Budget Directive 15, Standards for the Classification of Federal Data on Race and Ethnicity	Race identification, Ethnicity identification	A
Occupation Classification	Bureau of Labor Statistics, Standard Occupational Classification System	Job activity classification	A
Vital Statistics	Centers for Disease Control (CDC) National Center for Health Statistics	Birth and death records, medical records, interview surveys, physical exams, laboratory testing, marriages and divorces, fetal death	A
	Bureau of the Census Current Population Survey.	Primary source of information on the labor force characteristics of the U.S. population. The sample is scientifically selected to represent the civilian noninstitutional population.	C

Standard Type	Standard Name	Standard Content	Recommendation
Measurement	HL7 codes for Units, Versions 2.X + (derived from the ISO 2955-83 standard (withdrawn by ISO in 2001) and ANSI X3.50)	Common units of measure, such as Celsius or mg/ml, intended to be combined with a numeric value to accurately express a result.	A
	ISO 31, Quantities and units	Individual standards dealing with quantities in space and time, periodic phenomena, mechanics, heat, electricity and magnetism, electromagnetic radiation, chemistry, molecular physics, nuclear physics	A
Information Processing	FIPS 4-2, Representation of Calendar Date for Information Interchange	Means of representing calendar date to facilitate interchange of data among information systems	A
	ISO 8601, Numeric representation of dates and times	Formats for date and time	A
Education	United Nations Educational, Scientific, and Cultural Organization (UNESCO) International Standard Classification of Education (ISCED)	Designed to serve as an instrument suitable for assembling, compiling and presenting comparable indicators and statistics of education both within individual countries and internationally.	C
	National Center for Education Statistics (NCES)	Works with federal, state, and local education agencies and researchers, to develop a series of data handbooks to provide guidance on consistency in data definitions for education data.	C
Health-related Vocabulary/Coding Standards			
Basic Biology	Biological Pathways Exchange (BioPax)	Ontology for pathway information	B
	International Union of Biochemistry and Molecular Biology (IUBMB) and the International Union of Pure and Applied Chemistry (IUPAC)	Controlled vocabulary for nomenclature for biochemistry and molecular biology	A
Clinical	Common Terminology Criteria for Adverse Events v 3.0 (CTCAE)	Descriptive terminology for Adverse Event reporting.	A
	Current Procedural Terminology (CPT) 4	Coding for evaluation and management, anesthesia, surgery, radiology, pathology and laboratory, medicine	B

Standard Type	Standard Name	Standard Content	Recommendation
	Healthcare Common Procedure Coding System (HCPCS)	Healthcare procedures, equipment, and supplies (Level 1) - used for Medicare billing Classification (national level) of physician and non - physician patient care services (Level 2)	B
	International Classification of Diseases for Oncology (ICD-O-3)	Coding for diagnoses of neoplasms - both topography and morphology - includes tumor location, cell type, tumor type, aggressiveness grade	A
	International Classification of Diseases, Clinical Modification (ICD-9-CM)	Classifies diseases, conditions, symptoms, complaints/problems by diagnosis; supplementary classifications include health status, external causes of injury and poisoning, morphology of neoplasms, glossary of mental disorders, drug list numbers, industrial accidents and surgical, diagnostic, and therapeutic procedures	B
	International Statistical Classification of Diseases and Related Health Problems (ICD-10)	Collection, processing, classification, and presentation of mortality statistics	A
	Logical Observation Identifiers Names and Codes (LOINC)	Standard test names and codes, descriptive elements for other healthcare areas	A
	Medical Dictionary for Regulatory Activities (MedDRA)	Signs, symptoms, diseases, diagnoses, therapeutic indications, names and qualitative results, surgical and medical procedures, medical/social/family history, adverse event reporting	A
	Systematized Nomenclature of Human and Veterinary Medicine (SNOMED)	Findings/conclusions/assessments, procedures, body structures, function, organisms, substances, physical agents, occupations, social context/demographics, specimens, and other concepts	A
Genomics	Gene Ontology (GO)	Structured, controlled vocabularies describing gene products used for gene annotations.	A

Standard Type	Standard Name	Standard Content	Recommendation
	HUGO Gene Nomenclature Committee (HGNC)	Controlled vocabulary of gene names and symbols for human genes.	A
	Mammalian Phenotype Ontology (MP)	Standard vocabulary to describe phenotype data.	B
	The Microarray Gene Expression Data (MGED) Society	Comprised of MIAME, MAGE, and the MAGE ontology, a suite of standards for microarray users and developers including an object model, document exchange format, toolkit, and ontology.	A
	Mouse Anatomy (adult – MA, and development – EMAP)	Ontologies used to annotate gene products.	A
	Mouse Pathology (MPATH)	Ontology of terms used to annotate histopathology images.	C
	Sequence Ontology	Ontology of terms for use in annotation of biological sequences.	C
	Taxonomy	National Center for Biotechnology Information (NCBI) taxonomy of organism names represented in genetic databases.	A
Drug Identification	National Drug File Reference Terminology (NDF-RT)	Drug classes, active ingredients (chemical structure), mechanics of action, physiologic effect, pharmacokinetics, therapeutic intent, commercial/clinical drug identification	A
	RxNorm Clinical Drug Vocabulary	Ingredients, drug components, drug formulations, drug strength representation, drug name synonyms, dosage forms	A
Health-related Transaction Standards and Models			
	Digital Imaging and Communications in Medicine (DICOM)	Standard method for the transmission of medical images and their associated information.	A
Basic Biology	Systems Biology Markup Language (SBML)	XML exchange format for exchange of biochemical network models.	A
	CellML	XML-based language for describing and exchanging models of cellular and subcellular processes.	A

Standard Type	Standard Name	Standard Content	Recommendation
Clinical	American National Standards Institute (ANSI) X12	Insurance, claim payment, eligibility benefit inquiry, healthcare service coding, enrollment, encounter and claims	C
	American Society for Testing and Materials (ASTM) E1384-02a, Standard Guide for Content and Structure of Electronic Health Record	Patient identification (name, SSN, Birth Date, race, ethnicity), encounter, problem description, treatment, service type, testing (LOINC), appointment, provider information	C
	Clinical Data Interchange Standards Consortium (CDISC)	Clinical trials data - general data (study name, protocol name, measurement units), study metadata (code lists), administrative data, reference data (lab normal ranges), clinical data	A
	Health Level Seven (HL7)	Patient tracking, scheduling, orders, results, clinical observations, billing, medical records, patient referral, patient care	A
	North American Association of Central Cancer Registries, Inc. (NAACCR, Inc.)	Demographic, tumor and staging, treatment and follow-up	A
Genomics	Biomolecular Sequence Analysis (BSA)	Specification defines a data model and interface format for data exchange on biological sequences.	C
	Genomic Maps	Data model and structure for schema mapping for genomic maps.	C
	Macromolecular Structure (Mms)	Specification for a data model and interface for exchange of macromolecular structure information.	B
	PEDRo	Data model implemented in SQL and XML to support proteomics research	B
	Protein-Protein Interaction (PPI)	Data exchange format designed to bridge different formats of protein interaction databases.	B
	Tissue Microarray (TMA)	Data exchange specification for tissue microarray data.	A
Standards Review Bodies			

Standard Type	Standard Name	Standard Content	Recommendation
	Centers for Disease Control Public Health Information Network (PHIN)	Vocabulary and messaging standards; standards for data display and entry; standards for data transmission and management; implementation of applications and databases to support the adopted data standards.	
	Consolidated Health Informatics Initiative (CHI)	Portfolio of existing clinical vocabularies and messaging standards enabling federal agencies to build interoperable federal health data systems.	
	Food and Drug Administration, Center for Drug Evaluation and Research (CDER)	Compilation of standardized nomenclature monographs that have been reviewed and approved by the CDER Nomenclature Standards Committee (NSC)	
	National Council on Vital Health Statistics (NCVHS)	Advises the government on recommended standards for adoption in the health care sector	

Exhibit 1. Standard Type Table